OPONENTSKÝ POSUDEK NA DIPLOMOVOU PRÁCI

Název: Párové testy s chybějícími pozorováními
Autor: Kristýna Sionová

Shrnutí:

Diplomová práce Kristýny Sionové se zabývá testováním rovnosti středních hodnot dvou závislých veličin v normálním modelu za přítomnosti chybějících pozorování. Jedná se vlastně o zobecnění párového t-testu na chybějící data. Práce obsahuje přehled vybraných metod, většinou vyvinutých v 60. a 70. letech, a simulační studii chování jejich testových statistik za nulové hypotézy (hladina) a za alternativy (síla). Téma je to pěkné a zajímavé a Kristýna Sionová jej zpracovala kvalitně. Práce není příliš obsáhlá, má kolem 30 stránek textu, 20 stránek grafů (simulované síly testů) a několik stránek tabulek (simulované hladiny). Je v ní několik málo drobných faktických chyb (viz níže), ale nikoli závažných, a nalezl jsem jen pár překlepů v textu. Grafická a jazyková úroveň práce je slušná.

Jako slabší stránku práce bych viděl to, že se nesnaží vylepšit ani rozšířit studované metody, že se omezuje na výběry z normálního rozdělení, a že nedostatečně specifikuje předpoklady o chybějících pozorováních. V úvodu práce je řečeno doslova

> *. . . předpokládáme, že chybění jednotlivých pozorování je náhodné (způsobené například špatnou organizací experimentu), nikoli systematické (příkladem takové situace jsou například dvojice pozorování „body získané v příjmacím řízení" a „průměr známek z prvního ročníku VŠ" — známky z VŠ chybí nikoli pro náhodné studenty, ale pouze pro ty, kteří v příjmacím řízení dosáhli menšího než dostačujícího počtu bodů).*

Nicméně není zde vysvětleno, co to je přesně „náhodné chybění". To by vyžadovalo specifikaci rozdělení veličin, které se skrývají za „otazníky" (tj. těch chybějících) a specifikaci rozdělení otazníků. Příklad o výsledku přijímacího řízení a známkovém průměru není příliš přesvědčivý, protože tyto veličiny nemají sdružené normální rozdělení a testování rovnosti jejich středních hodnot není relevantní problém. Zdá se také, že v celé práci je počet chybějících hodnot každé veličiny považován za konstantu. Fakticky by jejich četnosti n_1 a n_2 měly být považovány za náhodné veličiny a diskuse o tom, do jaké míry jejich náhodnost ovlivňuje zkoumané metody by mohla být zajímavým příspěvkem.

Předloženou práci Kristýny Sionové celkově hodnotím jako velmi dobrou a doporučuji ji uznat za práci diplomovou.

Připomínky:

- 5_1 Překlep „příjmacím".
- 9_1 Překlep „popsného".
- 11^9 Tohle asi není $Z(0)$, $Z(0)$ by bylo rovno $\bar{X}^{(n)} - \bar{Y}^{(n)}$.
- 12^4 Anděl (Statistické metody, Matfyzpress 1998, str. 88) tuto metodu nazývá Satterthwaiteův test; Welchův test má podle Anděla trošku odlišnou aproximaci f.
- 14_2 Měla jste asi na mysli, že neexistuje *explicitní* vyjádření maximálně věrohodných odhadů kovariancí a rozptylů. Implicitní vyjádření zajisté existuje. Lin a Stivers byli asi i na svou dobu (1974) příliš zaujati odvozováním explicitních vzorců. Maximálně věrohodné odhady by šlo celkem snadno spočítat iterativní procedurou.
- 17_1 Test podle Welche: Lze Welchovu (či Satterthwaiteovu) aproximaci tímto způsobem aplikovat na tři výběry? Z čeho to plyne?
- Ke 4.1 a 4.2 (Bhojův test): Na Bhojově testu mne zaujalo, že vyžaduje, aby na obou veličinách alespoň něco chybělo. Pokud chybí hodnoty jen jedné veličiny, tak stačí zahodit alespoň

jednu hodnotu druhé veličiny — a ejhle, Bhojův test můžeme použít.

- Ke 4.1.1 a 4.2.1 (volba vhodného λ): Je jasné, že „vhodné" λ závisí zejména na korelaci ρ. Pokud jsou X a Y slabě korelovány, bude lepší na ně použít dvouvýběrový test, tj. malé λ, když budou silně korelovány, bude lepší dát větší váhu na párový test, tj. velké λ. Volit λ tak, aby byl průměrný výsledek přes všechny možné korelace co nejlepší, asi není optimální řešení. Co kdyby se prostě odhadla korelace ρ z úplných dvojic pozorování a našla se nejlepší lambda pro tuto odhadnutou korelaci? (Ve 4.2.1 by se odhadly i rozptyly σ_1^2 a σ_2^2.) Vznikl by tím jakýsi adaptivní test, který by se sám přizpůsoboval pozorovaným datům.
- Ke 4.2 (Bhojův test s nestejnými rozptyly): Odhad S' se velice okrádá, zvláště když $n_1 \ll n_2$. Proč neodhadnout rozptyl $\bar{X}^{(n_1)} - \bar{Y}^{(n_2)}$ přímo a nepoužít Welchovu aproximaci?
- Kapitola 5: Popis Mehtova a Gurlandova testu je značně neinformativní. V simulacích stejně tento test nebyl vyhodnocován, proč zde tedy tato kapitola je?
- 24^{16} Překlep „inervalového".
- 35^4 Překlep „nejlepěí".
- 35_5 „... porušení předpokladu testu podle Bhoje pro různé rozptyly." Je pravda, že test pro různé rozptyly předpokládá, že rozptyly nejsou stejné?
- 36_4 Překlep „ohdad".
- K závěrům: Zdá se mi, že pokud je korelace mezi oběma veličinami dost velká, pak si párový t-test na úplné páry vede velice dobře a žádná z těchto speciálních metod ho příliš nevylepší. Pokud je korelace malá, pak je zde ještě jedna jednoduchá alternativní metoda: udělat dvouvýběrový t-test na všechna pozorování. Vsadím se, že pro $\rho < 0.3$ by obyčejný dvouvýběrový t-test (třeba s Welchovou aproximací na nestejné rozptyly) přebil všechny metody na chybějící párová data.

Mgr. Michal Kulich, PhD.
KPMS MFF UK
4. září 2009